心理科学进展 2019, Vol. 27, No. 12, 2043–2051
Advances in Psychological Science DOI: 10.3724/SP.J.1042.2019.02043

语言和音乐层级结构的加工*

张晶晶 [1] 杨玉芳 [2]

([1] 南京师范大学心理学院, 南京 210097) ([2] 中国科学院心理研究所 行为科学重点实验室, 北京 100101)

摘 要 语言和音乐在呈现过程中, 小单元相互结合组成更大的单元, 最终形成层级结构。已有研究表明, 听者能够将连续的语流和音乐切分成层级结构, 并在大脑中形成层级表征。在感知基础之上, 听者还能将新出现的语言和音乐事件整合到层级结构之中, 形成连贯理解, 最终顺利地完成交流过程。未来研究应剖析边界线索在层级结构感知中的作用, 考察不同层级整合过程的影响因素, 进一步探索语言和音乐层级结构加工的关系。

关键词 语言; 音乐; 层级结构; 知觉; 整合

分类号 B842

语言和音乐作为人类最重要的两种交流系统, 在人类生存和发展中扮演了重要的作用。由于语言和音乐均涉及复杂而有意义的符号序列, 对两者关系的思考长久以来吸引了包括哲学家、诗人、语言学家、音乐学家等思想家的关注, 近些年来也在现代认知科学领域掀起了对比研究热潮(Patel, 2010; Patel & Morgan, 2017; 叶铮, 周晓林, 2006)。其中, 层级结构就是两者都具有的一种重要的组织特征。

语言在随时间展开过程中, 会借助韵律、句法、语义等语言线索, 将连续的语流或者文字切分成时间尺度不同的结构单元, 包括词语、短语、句子、段落乃至整个篇章。根据句法规则, 小的结构相互结合形成时间尺度更大的结构, 大结构再结合起来, 最终形成语言层级结构(Berwick, Friederici, Chomsky, & Bolhuis, 2013)。音乐同样也是如此。音乐在呈现过程中, 从来也不是永不停歇地进行, 而是根据音乐内容与音乐形象的需要, 通过不同的句读标志划分出不同级别的句法结构单元。音乐作品, 通过对这些句法结构单元的组合与排列构成一个完整的结构系统, 借以陈述音乐内容和表达音乐形象。这些结构单元根据时间尺度依次增大, 大致包括动机、乐节、乐句、乐段及其与它们同级的一些结构形式, 形成音乐层级结构(Lerdahl & Jackendoff, 1983)。

层级结构在语言、音乐理论中占据着重要的地位, 是语言和音乐进行结构组织的一种重要手段。那么对于信息接收者而言, 也就是当我们听别人说话或者欣赏音乐时, 能否感知到层级结构? 层级结构又会如何影响我们的理解与交流? 近些年来, 两个领域的研究者们采用相似的实验范式和实验手段, 对此展开了研究和探讨。本文将围绕这两个问题, 对语言和音乐领域的相关研究进行综述, 并在文章最后提出对未来研究的思考和展望。

1 层级结构的感知

我们在听别人说话或者欣赏音乐时, 不是将连续的声音信号知觉为一个整体, 也不是把每一个声音加工成一个个独立的个体。在实际加工过程中, 我们需要借助语言和音乐线索, 将连续的声音信号切分成不同的结构单元, 比如将语流切分成词语、短语和句子, 将音乐切分成动机、乐节和乐句。将连续的语言和音乐事件划分出不同的结构单元, 是完成理解和交流的一个基础过程。

1.1 语言层级结构的感知

对结构的感知能力, 即考察人们能否准确地划分出结构边界。除了句法和语义线索, 口语中

收稿日期: 2019-03-20
* 国家自然科学基金青年项目(31800914), 江苏省高校自然科学研究面上项目(18KJD190003)。
通信作者: 张晶晶, E-mail: jinger.zhang@outlook.com;
杨玉芳, E-mail: yangyf@psych.ac.cn

的结构边界通常还会伴随一些韵律线索，比如边界前音节延长、边界处音高重置、插入无声段等(van Petten & Bloom, 1999)。

研究者们首先考察了短语边界的加工。Steinhauer, Alter 和 Friederici (1999)给被试听觉呈现德语句子，试图采用 ERPs (event-related brain potentials)技术考察韵律边界如何影响句法歧义的消解。结果发现，对于包含暂时句法歧义的句子，合适的韵律切分有助于消解歧义，此时被试对这类句子的加工与正常句子没有差异；而对于句法正常的句子，不合适的韵律切分反而会造成“暂时句法歧义”，表现为相比正常句子，被试诱发出更大的 N400 和 P600。实验结果说明韵律边界能够即时地影响句法分析过程。这个研究另外一个重要的结果，是发现在每种实验条件下，句子的语调短语边界处都会诱发一种特殊的脑电成分。这种脑电成分的潜伏期在边界前音节出现 400~500 ms 左右，分布位于头皮中后部，是一种正向脑电波，研究者将之命名为中止正漂移(closure positive shift, CPS)。当去掉语调短语边界处的无声段后，CPS 仍然出现，表明 CPS 确实反映韵律边界加工，而不是由于被试感知到打断言语信号的停顿产生的。

为了进一步确定 CPS 的认知意义，Pannekamp, Toepel, Alter, Hahne 和 Friederici (2005)逐步去掉语句中的语义、句法和语音这些音段信息而只保留超音段的韵律信息，结果在语调短语边界处仍然诱发了 CPS，表明该成分的出现与音段信息无关，反映纯粹的韵律边界加工。此后，研究者们使用德语之外的语言材料，比如英语、汉语、韩语等(Hwang & Steinhauer, 2011; Li & Yang, 2009; Li & Yang, 2010; Pauker, Itzhak, Baum, & Steinhauer, 2011)，在语调短语边界处均发现了 CPS。

除了语调短语边界，人们能否加工其他不同层级的语言结构呢？Li 和 Yang (2009)使用现代汉语句子作为实验材料，用 ERPs 技术考察句子内部三个不同层级边界的加工，包括韵律词边界、韵律短语边界和语调短语边界，另外选取双字词中两个字之间的边界(音节边界)作为基线条件。结果发现，与基线条件相比，不仅语调短语边界处诱发了 CPS，在层级更低的韵律短语边界处也发现了 CPS，但是后者的潜伏期更短，波幅也更小。对于最低层级的韵律词边界，虽然比基线条件诱发了更大的正波，但是在统计上并没有达到显著水平。除了句子内部的边界，随后研究者还使用唐代诗歌作为实验材料，考察了更大尺度的语篇中韵律边界的加工，不仅包括句子内部的边界，还包括句子之间的小句边界和联句边界等(Li & Yang, 2010)。结果发现相比于基线条件，所有边界水平都能诱发 CPS，并且 CPS 的潜伏期随着边界层级升高而发生系统的变化。这些研究结果共同表明，在语言加工中，人们能够感知不同层级的语言结构，并在头脑中形成层级表征。

那么，我们大脑如何能同时处理这些时间尺度完全不同的语言结构，即人们加工语言层级结构的神经机制是什么？一些研究发现(Ding & Simon, 2012; Luo & Poeppel, 2007; Peelle, Gross, & Davis, 2012)，让被试聆听口头语言时，被试大脑皮层中神经元的振荡频率会逐渐接近音节的呈现频率，从而将连续语流切分成不同的音节。在此基础之上，Ding, Melloni, Zhang, Tian 和 Poeppel (2016)考察了人们如何在线切分字、词语、句子三个不同的层级结构。实验中，给被试连续播放四个字组成的句子，例如“冰雪融化绵羊吃草……”，每个字呈现 250 ms。如此一来，字、词语、句子的呈现频率分别是 4 Hz、2 Hz 和 1 Hz。对所有的刺激去除了边界处的韵律线索，因此被试只能借助句法和语义线索完成切分。对 MEG (magnetoencephalography)收集的数据做频谱分析，发现了频率分别是 4 Hz、2 Hz 和 1 Hz 的神经振荡。说明大脑节律会对不同层级的语言结构形成层级追踪，从而帮助人们将连续语流切分成时间尺度不同的语言层级结构。

综上所述，研究者们借助认知神经科学技术，考察了语言层级结构的实时加工过程。结果发现，人们能够借助音段和非音段线索，将连续语流划分成时间尺度不同的层级结构，并在大脑中形成层级表征。人们划分语言层级结构的神经机制，可能是通过神经振荡追随语言结构频率实现的。

1.2 音乐层级结构的感知

和语言类似，音乐边界也伴随一些结构与声学线索(Lerdahl & Jackendoff, 1983)。结构线索包括对称、平行、和声进行等因素。声学线索指的是边界附近音符在时长、音高、力度、音色等方面发生的变化，特别是边界前音符的延长和边界处无声段的插入。

研究者采用 ERPs 方法考察了乐句边界的感知过程(Knösche et al., 2005; Nan, Knösche, & Friederici, 2009; Nan, Knösche, Zysset, & Friederici, 2008; Neuhaus, Knösche, & Friederici, 2006; Silva et al., 2014)。这些研究的音乐材料包含两个乐句，原材料的两个乐句之间包含一个休止符作为有边界条件，将休止符用音符填充作为无边界条件进行对照。ERPs 结果发现，相对于无边界条件，在边界处关键音符(边界后音符)出现之后 500 ms 左右诱发了一个双侧分布的中后部正波。由于它与言语中语调短语边界处诱发的中止正漂移有相似的波幅和头皮分布(Steinhauer et al., 1999)，因此将它命名为音乐 CPS (Knösche et al., 2005)。不同的是，音乐 CPS 通常在边界后音符上记录得到，而语言 CPS 的记录位置是在边界前音节。

后续研究发现除了无声段，音乐 CPS 还对其他一些边界线索敏感，包括边界前音符长度与和声终止，因此 CPS 反映乐句边界的加工，而不仅仅是对停顿的感知(Neuhaus et al., 2006)。除此之外，对 MEG 数据的源分析(Knösche et al., 2005)和 fMRI 的研究(Nan et al., 2008)发现，在乐句边界处激活了涉及注意和记忆的加工脑区，说明 CPS 可能反映更高层次的认知加工，包括将刚刚结束的第一个乐句储存在记忆之中，以及将注意转换到下一个乐句的加工(Neuhaus et al., 2006)。

除了乐句之外，在乐句之下和乐句之上还存在其他的结构，人们能否加工不同音乐层级结构的边界？Zhang, Jiang, Zhou 和 Yang (2016)采用音乐家和非音乐家两组被试，选用自然音乐作为实验材料，每一段音乐材料中包含三种不同层级的边界，层级从低到高依次是乐节边界、乐句边界和乐段边界。实验中给两组被试呈现这些音乐片段，为了保证被试认真听音乐，在 20%的试次中被试需要完成一个简单的再认任务。ERPs 结果发现，音乐家在所有结构边界处都诱发了 CPS，并且 CPS 的波幅随着边界层级升高而增大，表明音乐家能够识别不同层级结构的边界，并且在大脑中对音乐层级结构形成层级表征。非音乐家只在最大边界处诱发 CPS，此外在所有边界处诱发了前部分布的负波。边界处的负波可能表明非音乐家以另外一种策略加工边界(Neuhaus et al., 2006)，但是由于条件之间负波没有差异，说明他们虽然可以识别边界但是并不能够加以区分。两组参与者之间的比较，可以看出层级边界表征能力依赖于音乐训练。

在聆听音乐时，我们的大脑是如何将音乐切分成层级结构的？已经有少量研究对此进行了探讨。Doelling 和 Poeppel (2015)给音乐家和非音乐家呈现真实音乐片段，每个音乐片段呈现 13 秒。这些音乐片段的播放速度有快有慢，比如速度快的片段每秒呈现 8 个拍子，速度慢的片段每秒只呈现 0.7 个拍子。实验中让被试完成音高判断任务，用 MEG 考察听者对音乐的在线加工过程。结果发现，和语言一样，听者的大脑节律也会对音乐拍子形成追踪，但是这种追踪过程受到音乐经验的调节。对于各种播放速度的音乐，音乐家的大脑中都会形成相应频率的神经振荡，比如 8 Hz 和 0.7 Hz 的振荡，并且这种追踪过程随着音乐训练年限的增长而增强；与此相对，非音乐家的大脑节律只会追随快速音乐，并不能对慢速音乐(1 秒呈现的音符少于 1 个)进行追踪。拍子按照一定强弱规律的循环往复就构成更高层级的节拍。Nozaradan, Peretz, Missal 和 Mouraux (2011)同时考察了大脑加工拍子和节拍的神经机制。实验中选取 8 名有音乐经验的被试，给被试呈现音高相同的纯音序列，每秒呈现 2.4 拍。要求被试听这个纯音序列，或者把序列想象成二拍子节拍类型或者三拍子节拍类型。由于拍子的频率是 2.4 Hz，因此二拍子节拍的频率是 1.2 Hz，三拍子节拍的频率是 0.8 Hz。对 EEG (electroencephalography)信号进行分析发现，三种条件下被试大脑中都出现了 2.4 Hz 的响应；更为重要的是，尽管声音刺激在三种条件下都是一样的，想象二拍子节拍条件下出现了 1.2 Hz 响应，想象三拍子节拍条件下出现了 0.8 Hz 响应。实验结果表明，我们的大脑可以同时追踪拍子和节拍组成的层级结构。

以上研究表明，除了乐句结构，听者还能实时感知乐句之下和乐句之上的其他结构，即能够感知音乐层级结构，并且在大脑中形成层级表征。对于音乐层级结构感知机制的研究，目前还停留在相对基础的探索阶段，之后可以进一步考察大脑节律是否能够追随更大时间尺度的音乐结构。

1.3 语言、音乐层级结构感知的比较

综上所述，在语言和音乐呈现过程中，人们均可以借助结构和韵律线索，识别出不同层级结

构的边界，从而完成对于连续声音信号的切分过程。在层级结构的感知过程中，两个领域共享类似的认知神经机制，这表现在以下两个方面。第一，在加工语言和音乐结构边界时，产生了相似的 ERP 成分，即头皮中后部分布的正向脑电波 CPS。第二，对于连续语流和音乐的切分过程，都是通过大脑节律追踪声音信号频率实现的。

除了相似性，语言和音乐层级结构的感知也表现出一定的差异。首先，虽然人们均能识别出语言和音乐层级结构，但是形成层级表征的条件存在差异。在语言领域，人们通过内隐学习即可对语言层级结构形成层级表征；但是对于音乐而言，对音乐层级结构形成层级表征需要借助专业的音乐训练。此外，虽然加工语言和音乐结构边界都诱发了 CPS，但是 CPS 的潜伏期存在差异，语言 CPS 在边界前音节处诱发，而音乐 CPS 在边界后音符处诱发。在加工不同层级的边界时，随着结构层级的增高，语言 CPS 主要表现为潜伏期的系统变化，而音乐 CPS 表现为波幅的逐渐增大。因此，语言和音乐 CPS 的认知意义、以及两个领域结构感知的认知和神经机制有待进一步比较和厘清。

2 新信息在不同层级结构中的整合

在感知层级结构的基础之上，我们还需要将新出现的语言或者音乐事件整合到不同层级结构的背景信息之中，如此才能形成连贯的表征和理解。那么，不同层级的整合过程是如何共同发生的？它们在加工进程、加工难度和加工脑区等方面具有相似特征，还是存在差异呢？

2.1 不同语言层级结构中的语义整合

研究者们主要采用语义违反范式，考察了语义在不同层级结构中的整合过程。对于加工进程，存在两种不同的观点。一阶段模型认为(Hagoort & van Berkum, 2007)，不同层级的语义整合是同时进行的，在我们接触到新信息时，这些不同层级的语义整合也就即时发生了。与此相对，二阶段模型认为(Cutler & Clifton, 1999; Kintsch, 1998)，不同层级的整合过程是先后进行的，低层级语义整合先发生，在此基础上才会进行高层级语义整合。那么，不同层级的整合过程到底是先后进行还是同时发生的？采用具有高时间分辨率的 ERPs 技术，研究者们对这个问题进行了实证研究。Zhou 等人(2010)使用汉语句子作为实验材料，用违反范式考察了句子之中两种层级的语义整合过程。实验中，合理条件的例句为“*小赵修理一张长椅*”。对这个句子进行修改，形成两种层级的语义违反。低层级的违反是量词–名词构成的名词短语的违反，比如“*小赵修理一台长椅*”；高层级的违反是动词–名词构成的动宾短语的违反，比如“*小赵修理一张信纸*”。结果发现，不同层级的违反都诱发了 N400 成分，也就是不同层级的语义整合都发生在 400 ms 左右的时间窗内。采用德语句子作为实验材料，研究者们得到了一致的结果(Zhang, Jiang, Saalbach, & Zhou, 2011)。这两项研究至少表明，在句子之内，不同层级短语的整合过程都是即时发生的。

一些研究者使用语篇作为实验材料，考察了更大时间尺度上的整合进程。实验中首先给被试听觉播放一个语篇背景，例如“*As agreed upon, Jane was to wake her sister and her brother at five o'clock in the morning. But the sister had already washed herself, and the brother had even got dressed*”。然后视觉(van Berkum, Hagoort, & Brown, 1999)或者听觉(van Berkum, Zwitserlood, Hagoort, & Brown, 2003)呈现一个句子，“*Jane told the brother that he was exceptionally* ***quick/slow***”，最后一个词是关键词。关键词与最后一个句子之间的整合是较低层级的句子整合，关键词与语篇背景的整合是高层级的语篇整合。不难看出，如果仅考虑最后一个句子，不论关键词是 quick 还是 slow 都是语义合理的。如果关键词是 slow 时被试察觉到违反，一定是被试将关键词整合到高层级的语篇背景之中了。结果发现，无论采用听觉或者视觉呈现最后一个句子，关键词与语篇违反时都诱发了更大的 N400，表明人们能够将新出现的单词整合到高层级的语篇背景之中。进一步的比较发现，语篇 N400 与句子 N400 在形态、潜伏期、头皮分布上都不存在差异，表明句子整合和语篇整合是同时发生的，并不存在先后之分。

不同层级的语义整合在加工难度上是否有区别仍然存在争议。一些实验结果支持没有差异。例如 van Berkum 等人(1999, 2003)发现，无论口语加工还是书面语加工，当新出现的词语与所在句子或者语篇语义违反时，都会诱发 N400 效应，并且 N400 在潜伏期、波幅、头皮分布上都没有区

别, 表明句子和语篇两种层级的语义整合并不存在难度差异。但是另外一些 ERPs 研究却发现高层级的语义整合过程难度更大。例如, Zhou 等人(2010)发现, 尽管不同层级的语义违反都诱发了相似的 N400 效应, 但是相比于低层级的名词短语违反, 更高层级的动宾短语违反还诱发了头皮后部分布的晚期正成分, 表明高层级语义整合在后期需要耗费更多的认知资源。

除了加工进程和加工难度, 研究者还用 fMRI 比较了不同语言层级结构的整合脑区。实验中, 研究者将这个故事分别按照词语、句子或者段落截开, 然后随机排列, 形成三种材料版本(Lerner, Honey, Silbert, & Hasson, 2011)。通过不同时间尺度的语义连贯性操作, 考察词语、句子、段落三种不同层级结构的整合脑区。除此之外, 将整个故事完全逆序播放作为基线条件。结果发现, 随着层级结构逐渐递增, 加工脑区从初级听觉皮层向顶叶皮层、额叶皮层这些更高级的脑区移动, 在加工脑区上也表现出层级分布。聋哑人对于手语的加工也得到了类似的结果。Inubushi 和 Sakai (2013)以日语手语作为实验材料, 发现聋哑人在加工日语词汇、句子、语篇时, 随着语言层级结构的增加, 前额皮层的激活区域和激活强度也会逐渐增大。

综上所述, 研究者们比较了新信息在不同语言层级结构中的整合过程。在加工进程上, 不同层级的语义整合都是即时发生的, 并不存在先后之分。在加工难度上是否不同仍然存在争议, 一部分研究支持没有差异, 另外一些研究认为更高层级的整合过程需要耗费更多的认知资源。在加工脑区上, 不同的大脑区域分别负责不同层级的语义整合。

2.2　不同音乐层级结构中的句法整合

与语言具有明确的语义不同, 音乐的意义通常是模糊不确定的, 但是音乐具有相对明确的句法规则, 即音符的组织和排列需要遵循一组特定的规则(马谐, 杨玉芳, 张秋月, 2016; 张晶晶, 杨玉芳, 2017; 周临舒, 蒋存梅, 杨玉芳, 2012)。因此研究者们通常从句法入手, 通过句法违反范式, 考察不同音乐层级结构中的句法整合。

研究者们首先考察了低层级音乐句法的整合过程。实验中给被试呈现和弦或者旋律序列作为启动刺激, 之后呈现目标刺激。操纵目标刺激与启动刺激在句法上的关系, 考察句法整合过程。ERPs 研究发现, 当目标和弦与背景信息句法违反时, 相比句法合理, 会在 100~350 ms 处诱发早期右前负波(early right anterior negativity, ERAN) (Guo & Koelsch, 2015, 2016; Jentschke, Friederici, & Koelsch, 2014; Koelsch, Gunter, Friederici, & Schröger, 2000; Koelsch, Vuust, & Friston, 2018), 或者右侧颞叶前部负波(early right anterior negativity, RATN) (Patel, Gibson, Ratner, Besson, & Holcomb, 1998)。这种前部负波反映早期的句法加工过程, 一般解释为对于句法违反的探测。除此之外, 还在晚期阶段诱发了 N5 (Koelsch et al., 2000; Poulin-Charronnat, Bigand, & Koelsch, 2006)或者 LPC (Lagrois, Peretz, & Zendel, 2018; Patel et al., 1998; Zendel, Lagrois, Robitaille, & Peretz, 2015), 反映和声的整合过程和对于音乐结构的再分析。MEG (Maess, Koelsch, Gunter, & Friederici, 2001)研究发现, 句法违法相比句法合理条件, 在布洛卡区以及右侧镜像区域出现了更多的激活。从这些研究可以看出, 听众能够顺利地将当前信息整合到背景信息之中。

上述研究中的音乐背景, 仅仅是一组较短的和弦或者旋律序列(例如通常包含 4 个或者 7 个和弦), 句法整合过程发生在乐句或者乐句之下水平。除了乐句和乐句之下的句法整合, 听众能否完成更高层级的句法整合过程? 最近的几项 EEG 研究对此进行了考察。Koelsch, Rohrmeier, Torrecuso 和 Jentschke (2013)选用的实验材料包含两个乐句, 最后一个和弦是目标和弦。保持目标和弦与第二个乐句不变, 操纵第一个乐句的调性, 使得目标和弦与更高层级的第一个乐句之间的句法依存关系合理或者不合理。结果发现, 与低层级句法违反一样(Jentschke et al., 2014; Koelsch et al., 2000), 高层级句法违反同样诱发了早期前部负波 ERAN 和晚期负成分 N5, 表明听众可以完成乐句之外的句法整合过程。采用类似的实验材料, 一些研究者发现乐句之外的句法整合受到结构复杂度的影响, 相比简单的线性结构, 嵌套结构的整合过程更加依赖于专业的音乐训练(Ma, Ding, Tao, & Yang, 2018a; Ma, Ding, Tao, & Yang, 2018b)。

由此可见, 人们在聆听音乐时, 既可以进行乐句之内的句法整合, 也可以进行乐句之外的句法整合, 那么这两种不同层级的句法整合过程是

如何共同进行的？Zhang, Zhou, Chang 和 Yang (2018)采用两个乐句作为实验材料，分别操纵最后一个目标和弦与第一个乐句和第二个乐句之间的句法合理性，同时考察了乐句之内、乐句之外两种层级的句法整合。结果发现，不同层级的句法整合都能发生。在时间进程上，只有乐句之内的句法违反在早期诱发 ERAN 效应，表明低层级的句法整合过程更早发生。在晚期阶段不同层级的句法整合在 N5 上存在交互作用，可能表明低层级句法整合存在优先性。

使用时间尺度更大的音乐材料(长度为 4 分 15 秒), Farbood, Heeger, Marcus, Hasson 和 Lerner (2015)进一步考察了不同层级音乐结构的整合脑区。实验中将这个音乐片段按照小节、乐句或者乐段截开再随机排列，考察被试在加工不同的音乐版本时脑区激活情况。结果发现，音乐结构连贯性系统地影响脑区激活，原版本相比随机排列版本激活更多的额叶、顶叶等高级皮层区域，表明听者至少可以完成乐段层面的音乐整合。

总而言之，越来越多的研究支持听者能够将新出现的音乐事件整合到不同的层级结构之中。一些研究对音乐中不同层级的整合过程进行了比较，发现低层级整合更早发生，并且在加工上可能具有优先性。不同层级的音乐整合在加工脑区上也表现出一定的层级性，随着音乐层级结构的逐渐递增，加工脑区逐渐从初级听觉皮层向更高级的脑区移动。

2.3 语言、音乐层级结构中整合过程的比较

综上所述，在感知层级结构的基础之上，人们能够将新出现的语言和音乐事件整合到不同层级的结构之中，从而形成连贯的理解，顺利地完成交流过程。在认知神经机制上，两个领域的层级整合过程表现出一定的相似性。首先，高层级整合与低层级整合会诱发类似的 ERP 成分，比如语言中的 N400 和音乐中的 N5。其次，在加工脑区上，随着整合层级的增高，加工脑区从感觉皮层向更高级的认知皮层移动，在脑区上均表现出层级分布。

语言和音乐层级结构中的整合过程也存在差异。首先，在语言领域，虽然一些研究发现高层级整合需要付出更多的认知努力，但是大量研究都表明不同层级的整合过程是共同发生的。对于音乐而言，高层级整合的心理现实性仍然有待进一步验证，早期的研究甚至认为音乐中只存在此时此刻的加工，更高层级的整合过程并不存在(Tillmann & Bigand, 2004)。这或许是因为音乐中缺乏精确的意义辅助加工。此外，在加工进程上，语言中不同层级的整合过程是同时发生的，但是在音乐中，低层级整合过程更早发生，具有一定的优先性。加工进程上的差异，可能也表明音乐中高层级结构中的整合难度更大。

3 总结与展望

现有研究采用类似的实验方法和实验范式，从切分和整合两个方面考察比较了语言和音乐层级结构的加工，研究结果对于理解语言和音乐加工的本质，以及人类更一般的交流机制起着重要的作用。尽管如此，还有很多问题需要探索。首先，语言和音乐层级结构的感知研究可以进一步深入。关于语言和音乐结构边界的研究虽然数量不少，但是大部分研究都局限在单独的语调短语边界和乐句边界，对于其他结构边界乃至层级结构边界的感知，以及层级边界感知的认知神经机制，仍然有待更多的研究进行探索和验证。另外，语言和音乐结构的划分伴随句法、意义、韵律等线索，那么在边界感知过程中，哪一种线索起着更重要的作用，或者每种线索分别具有什么作用，这些问题目前并不清楚。

其次，已有研究发现人们会将即将出现的语言和音乐事件整合到不同层级的背景信息之中，接下来我们可以进一步探讨不同层级的整合过程受到哪些因素影响。比如 Egidi 和 Caramazza (2016)发现，不同层级的语义整合受到实验任务的影响。当被试任务是被动的阅读理解时，局部整合占据优势，把任务换成主动的合理性判断时，高层级的整合过程更加重要。除此之外，不同层级的整合还可能受到被试知识经验的影响(Ma et al., 2018a; Ma et al., 2018b)。例如, Ma 等人(2018b)选取非音乐家、音乐爱好者、音乐家三组音乐训练程度依次提高的被试，分别给他们呈现包含嵌套结构的西方音乐序列，结果发现，只有音乐家才能完成高层级句法整合过程，说明至少对于中国被试而言，在加工西方音乐时，高层级的整合过程可能依赖于专业音乐训练。

最后，语言和音乐层级结构加工的关系有待深入探讨。现有研究虽然采用相似的实验范式和

实验方法, 考察了语言、音乐层级结构的感知和整合过程, 也发现了音乐和语言在这两个过程上存在一些类似的认知和神经机制, 但是目前并没有一个研究同时考察语言和音乐层级结构的加工, 因此对比研究亟需开展。另外, 加工语言和音乐层级结构时类似的认知和神经机制也需要进一步挖掘探究。比如, 加工语言和音乐层级结构边界都会诱发脑电成分 CPS, 但是一些研究发现两者 CPS 的潜伏期存在差异, 一些研究认为潜伏期差异并不存在(Glushko, Steinhauer, DePriest, & Koelsch, 2016); 此外, 对音乐 CPS 的 MEG 数据进行源分析(Knösche et al., 2005)和 fMRI 的研究(Nan et al., 2008)发现, 在乐句边界处激活了涉及注意和记忆的加工脑区, 说明音乐 CPS 可能反映更高层次的认知加工, 那么语言 CPS 是不是也是这样, 这些问题仍然有待回答。

参考文献

马谐, 杨玉芳, 张秋月. (2016). 音乐句法的加工. *科学通报, 61*(10), 1099–1111.

叶铮, 周晓林. (2006). 音乐之脑. *心理科学进展, 14*(5), 641–647.

张晶晶, 杨玉芳. (2017). 音乐句法加工的影响因素. *心理科学进展, 25*(11), 1823–1830.

周临舒, 蒋存梅, 杨玉芳. (2012). 音乐和语言句法认知的比较. *科学通报, 57*(28–29), 2674–2685.

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences, 17*(2), 89–98.

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. *The Neurocognition of Language*, 123–166.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158–164.

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences, 109*(29), 11854–11859.

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences, 112*(45), E6233–E6242.

Egidi, G., & Caramazza, A. (2016). Integration processes compared: Cortical differences for consistency evaluation and passive comprehension in local and global coherence. *Journal of Cognitive Neuroscience, 28*(10), 1568–1583.

Farbood, M. M., Heeger, D. J., Marcus, G., Hasson, U., & Lerner, Y. L. (2015). The neural processing of hierarchical structure in music and speech at different timescales. *Frontiers in Neuroscience, 9*, 157.

Glushko, A., Steinhauer, K., DePriest, J., & Koelsch, S. (2016). Neurophysiological correlates of musical and prosodic phrasing: Shared processing mechanisms and effects of musical expertise. *PloS One, 11*(5), e0155300.

Guo, S., & Koelsch, S. (2015). The effects of supervised learning on event-related potential correlates of music-syntactic processing. *Brain Research, 1626*, 232–246.

Guo, S., & Koelsch, S. (2016). Effects of veridical expectations on syntax processing in music: Event-related potential evidence. *Scientific Reports, 6*, 19064.

Hagoort, P., & van Berkum, J. J. A. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 801–811.

Hwang, H., & Steinhauer, K. (2011). Phrase length matters: The interplay between implicit prosody and syntax in Korean “garden path” sentences. *Journal of Cognitive Neuroscience, 23*(11), 3555–3575.

Inubushi, T., & Sakai, K. L. (2013). Functional and anatomical correlates of word-, sentence-, and discourse-level integration in sign language. *Frontiers in Human Neuroscience, 7*, 681.

Jentschke, S., Friederici, A. D., & Koelsch, S. (2014). Neural correlates of music-syntactic processing in two-year old children. *Developmental Cognitive Neuroscience, 9*, 200–208.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY, US: Cambridge university press.

Knösche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., & Friederici, A. D. (2005). Perception of phrase structure in music. *Human Brain Mapping, 24*(4), 259–273.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: “Nonmusicians” are musical. *Journal of Cognitive Neuroscience, 12*(3), 520–541.

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences, 110*(38), 15443–15448.

Koelsch, S., Vuust, P., & Friston, K. (2018). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences, 23*(1), 63–77.

Lagrois, M.-É., Peretz, I., & Zendel, B. R. (2018). Neurophysiological and behavioral differences between older and younger adults when processing violations of tonal structure in music. *Frontiers in Neuroscience, 12*, 54.

Lerdahl, F., & Jackendoff, R. S. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT press.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011).

Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience, 31*(8), 2906–2915.

Li, W., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience, 158*(4), 1416–1425.

Li, W., & Yang, Y. (2010). Perception of Chinese poem and its electrophysiological effects. *Neuroscience, 168*(3), 757–768.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001–1010.

Ma, X., Ding, N., Tao, Y., & Yang, Y. F. (2018a). Differences in neurocognitive mechanisms underlying the processing of center-embedded and non–embedded musical structures. *Frontiers in Human Neuroscience, 12*.

Ma, X., Ding, N., Tao, Y., & Yang, Y. F. (2018b). Syntactic complexity and musical proficiency modulate neural processing of non-native music. *Neuropsychologia, 121*, 164–174.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4*(5), 540–545.

Nan, Y., Knösche, T. R., & Friederici, A. D. (2009). Non-musicians' perception of phrase boundaries in music: A cross-cultural ERP study. *Biological Psychology, 82*(1), 70–81.

Nan, Y., Knösche, T. R., Zysset, S., & Friederici, A. D. (2008). Cross-cultural music phrase processing: An fMRI study. *Human Brain Mapping, 29*(3), 312–328.

Neuhaus, C., Knösche, T. R., & Friederici, A. D. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *Journal of Cognitive Neuroscience, 18*(3), 472–493.

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience, 31*(28), 10234–10240.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience, 17*(3), 407–421.

Patel, A. D. (2010). *Music, language, and the brain*. New York, NY: Oxford university press.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717–733.

Patel, A. D., & Morgan, E. (2017). Exploring cognitive relations between prediction in language and music. *Cognitive Science, 41*(s2), 303–320.

Pauker, E., Itzhak, I., Baum, S. R., & Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *Journal of Cognitive Neuroscience, 23*(10), 2731–2751.

Peelle, J. E., Gross, J., & Davis, M. H. (2012). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23*(6), 1378–1387.

Poulin-Charronnat, B., Bigand, E., & Koelsch, S. (2006). Processing of musical syntax tonic versus subdominant: An event-related potential study. *Journal of Cognitive Neuroscience, 18*(9), 1545–1554.

Silva, S., Branco, P., Barbosa, F., Marques-Teixeira, J., Petersson, K. M., & Castro, S. L. (2014). Musical phrase boundaries, wrap-up and the closure positive shift. *Brain Research, 1585*, 99–107.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 191–196.

Tillmann, B., & Bigand, E. (2004). The relative importance of local and global structures in music perception. *The Journal of Aesthetics and Art Criticism, 62*(2), 211–222.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657–671.

van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*(3), 701–718.

van Petten, C., & Bloom, P. (1999). Speech boundaries, syntax and the brain. *Nature Neuroscience, 2*(2), 103–104.

Zendel, B. R., Lagrois, M.-É., Robitaille, N., & Peretz, I. (2015). Attending to pitch information inhibits processing of pitch information: The curious case of amusia. *Journal of Neuroscience, 35*(9), 3815–3824.

Zhang, J. J., Jiang, C. M., Zhou, L. S., & Yang, Y. F. (2016). Perception of hierarchical boundaries in music and its modulation by expertise. *Neuropsychologia, 91*, 490–498.

Zhang, J. J., Zhou, X. F., Chang, R. H., & Yang, Y. F. (2018). Effects of global and local contexts on chord processing: An ERP study. *Neuropsychologia, 109*, 149–154.

Zhang, Y., Jiang, X. M., Saalbach, H., & Zhou, X. L. (2011). Multiple constraints on semantic integration in a hierarchical structure: ERP evidence from German. *Brain Research, 1410*, 89–100.

Zhou, X. L., Jiang, X. M., Ye, Z., Zhang, Y. X., Lou, K. Y., & Zhan, W. D. (2010). Semantic integration processes at different levels of syntactic hierarchy during sentence comprehension: An ERP study. *Neuropsychologia, 48*(6), 1551–1562.

Processing of hierarchical structures in language and music

ZHANG Jingjing[1]; YANG Yufang[2]

([1] *School of Psychology, Nanjing Normal University, Nanjing* 210097, *China*)

([2] *CAS Key Laboratory of Behavioral Science, Institute of psychology, Beijing* 100101, *China*)

Abstract: In language and music development, smaller elements combine into larger units, which eventually form hierarchical structures. Previous studies have found that listeners can segment continuous stream of speech and music into hierarchies and represent them in a hierarchical way. On the basis of perception, listeners can also integrate incoming language and music events into hierarchical structures to form a coherent understanding and ultimately complete the communication smoothly. Future studies should analyze the role of boundary clues in hierarchical structure perception, examine the influencing factors of integration processes at different time scales, and further explore the relationship between language and music hierarchical structure processing.

Key words: language; music; hierarchical structure; perception; integration